Using Neural Network to Enhance Assimilating Sea Surface Height Data into an Ocean Model

Vladimir Krasnopolsky, Carlos J. Lozano, Deanna Spindler, Ilya Rivin, and D.B. Rao

Abstract—A generic approach that allows extracting functional nonlinear dependencies and mappings between atmospheric or ocean state variables in a relatively simple form is presented. These dependencies and mappings between the 2- and 3-D fields of the prognostic and diagnostic variables are implicitly contained in the highly nonlinear coupled partial differential equations of an atmospheric or ocean dynamical model. They also are implicitly contained in the numerical model output. An approach based on using neural network techniques is developed here to extract the inherent nonlinear relationship between the sea surface height anomaly and the other dependent variables of an ocean model. Specifically, numerically generated grid point fields from the Real Time Ocean Forecast System (RT-OFS) model of NCEP (National Centers for Environmental Prediction) are used for training and validating this relationship. The accuracy of the NN emulation is evaluated over the entire domain of the NCEP's RT-OFS. The differences in the accuracies of the technique in the coastal areas and in the deep water are discussed. Accurate determination of such relationships is an important first step to enhance the assimilation of the sea surface height measurements into an ocean model by propagating the signal to other dependent variables through the depth of the model.

I. INTRODUCTION

The output of any complex geophysical numerical model, like models for numerical weather prediction, climate and ocean simulations, contains a great amount of data in the form of 2- and 3-D high resolution numerical fields of prognostic and diagnostic variables. This output contains, in an implicit form, the highly complex functional dependencies and mappings between the dependent state variables of the model. These relationships are governed by the physics and dynamics of the numerical model that has been used for the simulations. A clear understanding of the underlying nonlinear dependencies is a matter of great scientific interest and practical importance.

For example, when 2-D observations like surface wind, surface currents, or sea surface elevation are assimilated in the atmospheric or oceanic data assimilation system (DAS), the impact of this data in the DAS is localized at the level of their assimilation because there is usually no explicit mechanism in the DAS to propagate the impact of these data to other vertical levels and to other variables. Usually this propagation occurs later, during the time integration of the model, in accordance with dependencies determined by the model physics. Recently several attempts have been made to extract simplified linear dependencies of such a kind from observed data [1] or model simulations [2] for the use in an ocean DAS. However, these simplified and generalized linear dependencies that are often derived from local data sets do not properly represent the complicated nonlinear relationship between the model variables. If we were able to extract/emulate these dependencies in a simple, but not overly simplified, and yet adequately nonlinear analytical form, they could be used in the DAS to facilitate a more effective 3-D assimilation of the 2-D surface data. These analytical functions and mappings could also be used for efficient model output compression, archiving, and dissemination, and for sensitivity studies and error analysis.

Existence of a generic technique that would allow extracting these nonlinear functions and mappings in a compact analytical form would greatly facilitate the use of model output for qualitative and quantitative studies. It is only recently that steps are being taken to use the NN technique to accomplish this objective. Here we introduce a generic NN technique using a particular application to the NN emulation for sea surface height. This new and generic NN application requires a reasonable quality of the Jacobian of the NN emulation. The Jacobian analysis and an ensemble approach to improve the quality of the NN emulation and NN Jacobian are presented in [3].

II. SSH MAPPING AND ITS NN EMULATION

A. Ocean Model

Sea surface height (SSH), η, is one of the prognostic variables in ocean circulation models. The particular ocean model that was used in this study is the HYbrid Coordinate Ocean Model (HYCOM). This model is a primitive equation model that uses a generalized hybrid coordinate (isopycnal/terrain following (σ)/z-level) in the vertical (see [4] for details). The hybrid coordinate extends the geographic range of applicability of traditional isopycnal coordinate (coordinates that follow the selected levels of constant water density) circulation models, toward shallow coastal seas and unstratified parts of the ocean. The vertical coordinate evaluation for HYCOM is discussed in [5]. The particular version of HYCOM used in this study has a domain that covers the entire Atlantic Ocean with an average horizontal resolution of ⅓°; and 25 vertical levels.

Fig.1. RMSE (in cm) on the test set for NNs (2) with different complexity (the number of hidden neurons).

V. Krasnopolsky, C. J. Lozano, D. Spindler, I. Rivin, and D.B. Rao are with the National Centers for Environmental Prediction, Camp Springs, MD 20746 USA (corresponding author V. Krasnopolsky phone: 301-763-8000 x 7262; fax: 301-763-8545; e-mail: Vladimir.Krasnopolsky@ noaa.gov).
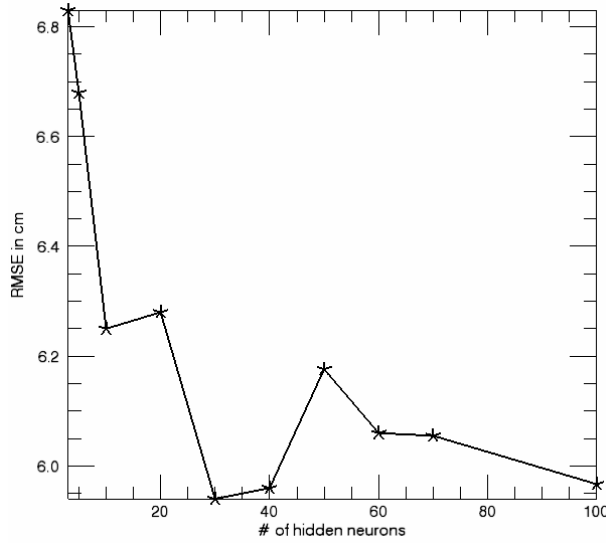
B. Developing the Mapping

Numerical models for the oceans and atmosphere are dissipative dynamical systems. In the phase space the model trajectories orbit attractors, usually of low dimension; or nearly satisfying reduced physics equations. A suitable choice of input variables is required to discover the desired relationship from the model states, without further prejudice. In a layered ocean model the SSH signal depends in part on the disposition of the layers in a vertical column. The reduced physics is the hydrostatic equation and the observation that in general pressure compensates at depth. Since the reduced physics model has mainly a 1-D vertical structure, we assumed, in this initial attempt, that SSH, or η, at a particular model grid point (i.e., at a particular horizontal location lat/lon) depends only on the vector of state variables X at the same horizontal location. Therefore, this dependence (a target mapping) can be written as

$$(1) \qquad \eta = \phi(X)$$

where ϕ is a nonlinear continuous function and X is a vector that represents a complete set of state variables, which determines SSH. In this particular case the vector X was selected as $X = \{I, \theta, z_{mix}\}$, where I is the vector of interfaces (vertical coordinates used in HYCOM), θ is the vector of potential temperature, and z_{mix} is the depth of the ocean mixed layer (a total of 50 variables). This set of variables represents (or used as proxy for) the physics of the deep ocean. Therefore, the mapping (1) with this particular selection of components for the vector X will not be applicable in costal areas (depth less then $250 - 500$ m). For the coastal areas a different set of state variables should be selected. All statistics presented below in this section were calculated using the test set where coastal areas were excluded.

The NN technique is applied to derive an analytical NN emulation for the relationship between model state variables, X, and SSH, or η,

$$(2) \qquad \eta_{NN} = \phi_{NN}(X)$$

using the simulated model fields which are treated as error free data. These fields are simulated by the model without data assimilation. A simulation that covers almost two years (from Julian day 303, 2002 to 291, 2004) was used to create training, validation and test data sets. The periods covered by these data sets and their sizes are shown in Table 1. Each data set consists of records $\{\eta_i, X_i\}_{i = 1,...,N}$ collocated in space and time and uniformly distributed over the model domain.

C. Evaluating the Mapping Accuracy

As mentioned earlier, the accuracy of the NN emulation is evaluated over the model domain (with coastal areas excluded) using the test set described in Table 1. Fig.1 shows improvement in the accuracy of the NN emulation (2) (RMSE) with the increase of the complexity (the number of hidden neurons) of the emulating NN. All trained NNs have 50 inputs and one output in accordance with the dimensionalities of the target mapping (1). The number of hidden neurons, k, was varied form 3 to 100. There is no significant and stable improvement in the RMSE after the number of hidden neurons, k, reaches values of $5 - 10$.

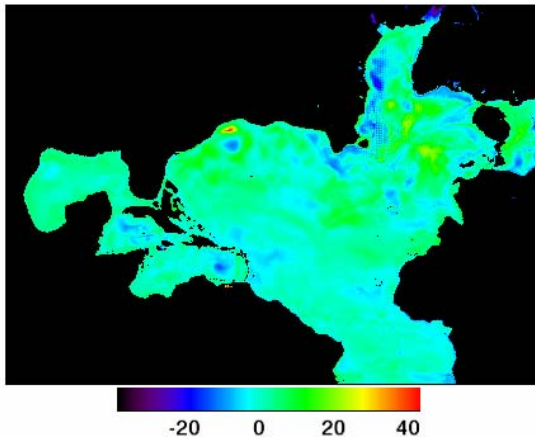TABLE 1.   PERIODS COVERED BY TRAINING, VALIDATION AND TEST DATA SETS AND THEIR SIZES.

Set	Beginning Date (Julian day, year)	End Date (Julian day, year)	Size, N (number of profiles)
Training	303, 2002	52, 2004	563,259
Validation	303, 2002	52, 2004	563,259
Test	53, 2004	291, 2004	563,259

Therefore, to limit the NN complexity and improve its interpolation abilities, we use only NNs with $k \leq 10$ in the following investigation.

In the next test applied to the NN emulation, the last day of the entire simulation (291, 2004) was selected. This day is separated by the time interval of about eight month from the last day of simulation used for training and validation (52, 2004). Fields generated by the model at 00Z were used to create inputs, X, for the NN emulation (2). Then the NN emulation (2) was applied over the entire domain (with coastal areas excluded) to generate 2-D field of η_{NN}. This field was compared with the corresponding field of SSH, η, generated by the model. The difference between two fields

Fig.2 The difference between SSH field η_{NN} emulated by NN (2) with 5 hidden neurons.

is shown in Fig.2. With the exception of several spots (most of them are still close to the coastal areas) the difference does not exceed ± 10 cm. The accuracy of the NN

[5] E. P. Chassignet, L. T. Smith Jr., G. R. Halliwell Jr., and R. Bleck, "North Atlantic simulations with the Hybrid Coordinate Ocean Model (HYCOM): Impact of the vertical coordinate choice, reference pressure, and thermobaricity", *J. Phys.Oceanogr.,* 33**,** pp. 2504–2526, 2003.

emulation over the entire domain shown in Fig. 2 is considered to be satisfactory: the bias is about 1cm and RMSE is about 4.7 cm. This conclusion that the accuracy of the NN emulation (2) is adequate is based on the fact that the NN emulation will be used in the DAS using satellite measurements of SSH that have an accuracy of order of 5 cm (or less).

The accuracy of the NN emulation may be improved using a NN ensemble approach. An additional improvement of accuracy of about 10 – 15 % can be achieved for this application [3]. The use of the NN emulation in the DAS is conditioned by the quality of the NN Jacobian. The accuracy of the NN Jacobian and the possibility to improve this accuracy using NN ensembles are discussed in [3].

III. CONCLUSION

In this paper we presented a generic NN approach that can be used to extract functional nonlinear dependencies and mappings between atmospheric and/or oceanic state variables using outputs from a numerical ocean model. These functions and mappings can be extracted in a closed, compact analytical form of NN emulations. Here we considered one particular application of this approach: using NN emulation of a mapping between a surface and subsurface parameters to enhance the data assimilation of a surface parameter (SSH) in the ocean DAS.

The availability of such NN emulations would greatly facilitate qualitative and quantitative studies of model output in general. These analytical functions and mappings could also be used for efficient model output compression, archiving, dissemination, and for sensitivity studies, and error analysis.

REFERENCES

[1] S. Guinehut, P.Y. Le Traon, G. Larnicol, S. Philipps, "Combining Argo and remote-sensing data to estimate the ocean three-dimensional temperature fields – a first approach based on simulated data", *J Marine Systems*, 46, pp. 85-98, 2004

[2] G. L. Mellor and T. Ezer, "A Gulf Stream model and an altimetry assimilation scheme", *J Geophys. Res.*, 96, pp. 8779-8795, 1991

[3] V. M. Krasnopolsky, "Calculation of a stable NN Jacobian using an ensemble approach", *Proceedings of the IJCNN2006, July 16-21, 2006*, Vancouver, BC, Canada, 2006, in press

[4] R. Bleck, "An oceanic general circulation model framed in hybrid isopycnic-cartesian coordinates", *Ocean Modell.,* 4**,** pp. 55–88, 2002.